First Order Asymptotic Theory

Elvezio Ronchetti

Research Center for Statistics
and
Geneva School of Economics and Management
University of Geneva, Switzerland

Elvezio.Ronchetti@unige.ch

http://www.unige.ch/ses/metri/ronchetti/

♦ Outline

- **Motivation**
- **Asymptotic Theory**
- $M-$**estimators**
  - Consistency
  - Asymptotic Normality
- **Berry-Esseen Bound**
- **Conclusions**

♦ Motivation

General Problem

Tail probabilities $P[T_n > t]$ are needed to carry out statistical inference (tests and confidence intervals).

Unless T_n has a simple form (e.g. linear)

and/or

the underlying distribution of the observations has a particular form (e.g. normal), tail probabilities cannot be computed exactly.

$\longrightarrow$ rely on asymptotic approximations

- A good asymptotic theory is one that works when $n = 1$.

- The purpose of asymptotic theory in statistics is simple : to provide usuable approximations before passage to the limit.

J.W. Tukey

Stirling's Formula

$$n! \sim \sqrt{2\pi}\ e^{-n}\ n^{n+1/2}$$

n	$n!$	Stir. appr.	% rel. err.
1	1	.92	8.0
2	2	1.92	4.0
3	6	5.84	2.7
4	24	23.51	2.0
5	120	118.02	1.6

♦ Asymptotic Theory

Central Limit Theorem

$X_1, \ldots, X_n \quad iid$
$E[X_i] = \mu \ , var[X_i] = \sigma^2 < \infty$

$\bar{X}_n = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$

$$F_n(t) := P\left[\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} < t\right]$$

$$CLT : F_n(t) \to \Phi(t) \quad n \to \infty$$

Nonlinear statistics:

Linearize statistic
by Taylor or von Mises expansion

♦ M-estimators

$$Z_1, \ldots, Z_n \qquad \text{iid} \qquad Z_i \sim F$$

$T(F) : E_F\left[\psi(Z_i; T(F))\right] = 0$
$F = \hat{F}$ empirical distribution :
$T_n = T(\hat{F}) : \sum_{i=1}^{n} \psi(Z_i; T_n) = 0$

M-estimator
Huber(1964), *Ann. Math. Stat.*

Consistency of M-estimators

Following e.g.:
Newey & McFadden(1994)
Handbook of Econometrics, Ch. 36

Θ: metric space (parameter space)
Q: $\Theta \to \mathbb{R}$
Q_n: sequence of random functions on Θ
$T_n = \arg\ \max_{\theta\in\Theta} Q_n(\theta)$

Assume:

(C1) uniform convergence
$\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{P} 0$

(C2) well-separated point of maximum
There exists a point θ_0 such that
$Q(\theta_0) > \sup_{\theta\in\Theta\setminus G} Q(\theta)$
for every open set G that contains θ_0.

Then, $T_n \xrightarrow{P} \theta_0$ as $n \longrightarrow \infty$.

In our case:

- $Z_1, \ldots, Z_n$ iid $Z_i \sim F = F_{\theta_0}$
- $Q(\theta) = E_F\left[\rho(Z_i; \theta)\right]$
- $Q_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \rho(Z_i; \theta)$
- T_n is an $M-$estimator
  with $\psi(z; \theta) = \frac{\partial}{\partial \theta} \rho(z; \theta)$
- (C1): law of large numbers required to hold uniformly in θ

Proof:

$\epsilon > 0$
$G = \{\theta \in \Theta | d(\theta, \theta_0) < \epsilon\}$, an open $\epsilon-$ball centered at θ_0.

To show: $P(T_n \notin G) \longrightarrow 0$.

Define: $\eta = Q(\theta_0) - \sup_{\theta \in \Theta \backslash G} Q(\theta)$.
By (C2), $\eta > 0$.

Then
$\{T_n \notin G\} \Rightarrow A_n = \{Q(T_n) \leq Q(\theta_0) - \eta\}$, thus
$P(T_n \notin G) \leq P(A_n)$, and
we show that $P(A_n) \longrightarrow 0$.

Define:

$$
\begin{aligned}
B_n &= \{|Q_n(T_n) - Q(T_n)| > \eta/2\} \\
C_n &= \{|Q_n(\theta_0) - Q(\theta_0)| > \eta/2\} \\
D_n &= \{\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| > \eta/2\}
\end{aligned}
$$

$B_n \subseteq D_n,\ C_n \subseteq D_n$
$\Rightarrow P(B_n) \leq P(D_n),\ P(C_n) \leq P(D_n)$, and
$A_n \cap B_n^c \cap C_n^c = \emptyset$

(If A_n, B_n^c, C_n^c should occur simultaneously, we must have
$Q_n(T_n) \leq Q(\theta_0) - \eta/2 < Q_n(\theta_0)$,
which contradicts that T_n maximizes Q_n.)

Then

$$
\begin{aligned}
P(A_n) &\leq P(A_n \cup (B_n \cup C_n)) \\
&= P(A_n \cap (B_n \cup C_n)^c) + P(B_n \cup C_n) \\
&= P(A_n \cap B_n^c \cap C_n^c) + P(B_n \cup C_n) \\
&= P(\emptyset) + P(B_n \cup C_n) \\
&\leq P(B_n) + P(C_n) \\
&\leq 2P(D_n)
\end{aligned}
$$

By (C1), $P(D_n) \longrightarrow 0$ and this proves the result.

Fisher consistency of M-estimators

Fisher consistency: $T(F_{\theta_0}) = \theta_0$

Sufficient condition for Fisher consistency of $M-$estimator: $E_{F_\theta}[\psi(Z_i;\theta)] = 0$

If T is Fisher consistent and continuous (with respect to the weak topology), then T is consistent.

Proof:

$$\widehat{F} \longrightarrow F_{\theta_0} \Rightarrow T_n = T(\widehat{F}) \xrightarrow{P} T(F_{\theta_0}) = \theta_0.$$

Asymptotic Normality of M-estimators

$Z_1, \ldots, Z_n$ iid $Z_i \sim F$ Huber (1967)

(A1) $T(F)$ is an interior point of the parameter space and an isolated root of $E_F\left[\psi(Z_i; T(F))\right] = 0$.

(A2) $\dot{\psi}(z;t) = \frac{\partial}{\partial t}\psi(z;t)$ is continuous at $t = T(F)$ uniformly in z.

(A3) $E_F\left[\psi(Z_i; T(F)) \cdot \psi^t(Z_i; T(F))\right] = Q(T,F) < \infty$.

(A4) $E_F\left[-\dot{\psi}(Z_i; T(F))\right] = M(T,F) < \infty$ and nonsingular.

Then, if $T_n \xrightarrow{P} T(F)$ as $n \longrightarrow \infty$,

$$\sqrt{n}\,(T_n - T(F)) \underset{n\to\infty}{\xrightarrow{\mathcal{D}}} \mathcal{N}(0, V(T,F)),$$

where $V(T,F) = M^{-1}(T,F)Q(T,F)M^{-t}(T,F)$.

Proof:

$\theta_\circ := T(F)$

- Taylor expansion of the estimating equation

$$
\begin{aligned}
0 &= \sum_{i=1}^{n} \psi(Z_i; T_n) \\
&= \sum_{i=1}^{n} \psi(Z_i; \theta_\circ) + \sum_{i=1}^{n} \dot{\psi}(Z_i; \theta^*) \cdot (T_n - \theta_\circ),
\end{aligned}
$$

where $\|\theta^* - \theta_\circ\| \leq \|T_n - \theta_\circ\|$

$$
\Rightarrow \sqrt{n}(T_n - \theta_\circ) = \left\{ \frac{1}{n} \sum_{i=1}^{n} -\dot{\psi}(Z_i; \theta^*) \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(Z_i; \theta_\circ)
$$

- Central Limit Theorem, Law of Large Numbers and Slutski Theorem

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(Z_i;\theta_\circ)\ \underset{n\to\infty}{\xrightarrow{\mathcal{D}}}\ \mathcal{N}(0,Q)$$

$$\frac{1}{n}\sum_{i=1}^{n}-\dot{\psi}(Z_i,\theta^*) =$$

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}-\dot{\psi}(Z_i;\theta_\circ)}_{\substack{n\to\infty\quad\downarrow\quad P\\ M}} +$$

$$\underbrace{-\frac{1}{n}\sum_{i=1}^{n}\left[\dot{\psi}(Z_i;\theta^*)-\dot{\psi}(Z_i;\theta_\circ)\right]}_{\substack{\downarrow\quad P\\ 0}}$$

(consistency and (A2))

Special case :

$F = F_\theta$ with density f_θ

$$\psi(z;\theta) = \frac{\partial \log f_\theta(z)}{\partial\theta}$$

$\Rightarrow T_n$: Maximum Likelihood Estimator.

In this case :

$$Q(T, F_\theta) = M(T, F_\theta) = J(\theta)$$

(Fisher information)

$\Rightarrow V(T, F_\theta) = J^{-1}(\theta)$

Properties

- Influence function:

$$IF(z;\psi,F) = M(\psi,F)^{-1}\psi(z;T(F))$$

- To any asymptotically normal estimator, there exists an asymptotically equivalent M-estimator.

  Model F_θ, a Fisher consistent estimator U for θ (i.e. $U(F_\theta) = \theta$) with $IF(z;U,F_\theta)$ $\longrightarrow$ M-est. defined by the M-functional T^M

$$E_F\Big[\psi(Z;T^M)\Big] = 0\,,$$

  where $\psi(z;\theta) = IF(z;U,F_\theta)$ has the same asymptotic distribution as U.
  Hampel, Ronchetti, Rousseeuw, Stahel(1986), p. 231.

Proof:

- T^M is Fisher consistent since by construction

$$E_{F_\theta}[\psi(Z;T^M(F_\theta))] = E_{F_\theta}[IF(Z;U,F_{T^M(F_\theta)})] = 0$$

$$\longrightarrow T^M(F_\theta) = \theta$$

-

$$
\begin{aligned}
& IF(z;T^M,F_\theta) \\
& = \Big(-E_{F_\theta}\Big[\frac{\partial\psi(Z;t)}{\partial t}\Big|_{t=T^M(F_\theta)}\Big]\Big)^{-1}\psi(z;T^M(F_\theta)) \\
& = \Big(-E_{F_\theta}\Big[\frac{\partial\psi(Z;\theta)}{\partial\theta}\Big]\Big)^{-1}\psi(z;\theta) \\
& = \Big(-E_{F_\theta}\Big[\frac{\partial IF(z;U,F_\theta)}{\partial\theta}\Big]\Big)^{-1} IF(z;U,F_\theta)
\end{aligned}
$$

- First term of von Mises expansion:

$$
\begin{aligned}
\tilde\theta - \theta &= U(F_{\tilde\theta}) - U(F_\theta) \\
&= \int IF(z;U,F_\theta) f_{\tilde\theta}(z)dz + o(\|\tilde\theta - \theta\|)
\end{aligned}
$$

Derivative with respect to $\tilde{\theta}$ at $\tilde{\theta} = \theta$:

$$
\begin{aligned}
I &= \left.\frac{\partial U(F_{\tilde{\theta}})}{\partial \tilde{\theta}}\right|_{\tilde{\theta}=\theta} = \int IF(z; U, F_\theta) \frac{\partial f_\theta(z)}{\partial \theta} dz \\
&= -\int \frac{\partial IF(z; U, F_\theta)}{\partial \theta} f_\theta(z) dz \\
&= -E_{F_\theta}\left[\frac{\partial IF}{\partial \theta}\right],
\end{aligned}
$$

where I is the identity matrix.

$\longrightarrow IF(z; T^M, F_\theta) = IF(z; U, F_\theta)$

$\longrightarrow$ M-estimator defined by ψ and U have the same asymptotic properties

Questions

(i) How good is the normal approx. ?

- Estimate error in CLT
- Improve this by providing asymptotic expansions

(ii) Simple as. approx. of $F_n(\cdot)$ which works well for small n

(iii) As $t \to \infty$: $\begin{matrix} F_n(t) \to 1 \\ \Phi(t) \to 1 \end{matrix}$ $\Rightarrow$ CLT ?

would like e.g. $\frac{1-F_n(t)}{1-\Phi(t)} \to 1 \begin{matrix} n \to \infty \\ t \to \infty \end{matrix}$

(Large deviations)

n	t	P	$\widehat{P}$	% $rel.error$
10	15	$6.98 \cdot 10^{-2}$	$5.69 \cdot 10^{-2}$	18
	20	$4.99 \cdot 10^{-3}$	$7.83 \cdot 10^{-4}$	84
	25	$2.21 \cdot 10^{-4}$	$1.05 \cdot 10^{-6}$	99
100	125	$9.38 \cdot 10^{-3}$	$6.21 \cdot 10^{-3}$	34
	150	$65.92 \cdot 10^{-6}$	$2.87 \cdot 10^{-7}$	95
	175	$2.78 \cdot 10^{-10}$	$3.19 \cdot 10^{-14}$	99
500	550	$1.15 \cdot 10^{-2}$	$1.27 \cdot 10^{-2}$	10
	600	$1.23 \cdot 10^{-5}$	$3.87 \cdot 10^{-6}$	68
	625	$1.01 \cdot 10^{-7}$	$1.13 \cdot 10^{-8}$	89

Exact tail probabilities $P = P[T_n > t]$,
normal approx. $\widehat{P}$,
relative error $\mid P - \widehat{P} \mid / P$,
statistic $T_n = \frac{1}{2}\sum_{i=1}^{n}(X_i^2 + Y_i^2)$,
X_i and Y_i iid $N(0,1)$,
exact: $T_n \sim \frac{1}{2}\chi^2_{2n}$.

Unfortunately the asymptotic (normal) distribution can be a poor approximation of tail areas especially for moderate to small sample sizes or far out in the tails.

This is exactly the region of interest for constructing confidence intervals and tests.

♦ Berry-Esseen Bound

Berry (1941); Esseen (1942)

If $E \mid X_i \mid^3 = \rho < \infty$, then

$$\mid F_n(t) - \Phi(t) \mid \leq \frac{3\rho}{\sigma^3 \sqrt{n}} \qquad \forall t, n$$

Remarks

- Bound depends only on the first three moments

- Can replace factor 3 with better upper bound (0.7975)

- Can be generalized to variables without third moment

- Berry - Esseen bound usually intolerable except for large samples.

Some References

For $U-$ statistics:

Bickel (1974), *Ann. Stat*
Helmers (1977), *Ann. Prob.*
Bjerve (1977), *Ann. Stat.*
Callaert & Janssen (1978), *Ann. Stat.*

For symmetric fct. of iid rv:

van Zwet (1984), *Zeit. W'keit. & Vew. Geb.*

⧫ Conclusions

- The asymptotic theory is a useful tool to understand the behavior of an estimator or test statistic.

- However, it can be unreliable as an approximation for moderate to small sample sizes and when approximations in the tails of the distribution are required.